HHS Public Access

Author manuscript

Sch Psychol. Author manuscript; available in PMC 2020 November 01.

Published in final edited form as:

Sch Psychol. 2019 November ; 34(6): 591–602. doi:10.1037/spq0000337.

Universal Screening in Middle and High Schools: Who Falls Through the Cracks?

Samantha M. Margherio, **Steven W. Evans**, **Julie Sarno Owens**

Ohio University

Abstract

Navigating academic demands in middle and high school may be particularly challenging for youth experiencing emotional and behavioral difficulties, and screening practices are a necessary first step in identifying youth in need of services. The goal of this study was to inform efficient universal screening practices in secondary schools by comparing three common methods of screening including self- and teacher-report using the Strengths and Difficulties Questionnaire (SDQ) and grade point average (GPA) to indicate moderate and high levels of risk. Participants were 2,350 secondary students (47% female; 94% non-Hispanic white) and their teachers in one socioeconomically-diverse school district. Using the moderate-risk criteria, there was low agreement among methods and almost half the sample (46%) was identified as at-risk by at least one of the three methods, indicating that responding to the needs of students who cross this liberal cutoff may not be feasible given school resources. Using the high-risk criteria reduced the agreement among screening measures. Self-report identified the most students as high-risk and was more sensitive to teacher-rated concerns than the converse. If students were uniquely identified by teacher-report, teachers rated them at high levels of risk, and predominant concerns were related to conduct problems. GPA captured few students with self- or teacher-rated internalizing or externalizing difficulties. Implications for universal screening within secondary schools are discussed.

Keywords

Universal screening; secondary schools; Strengths and Difficulties Questionnaire

Middle and high school represent potentially daunting transitional periods for adolescents with social, emotional, and behavioral (SEB) difficulties. Adjusting to increased academic demands, importance of peer relations, and opportunities to engage in risky behaviors prevalent in adolescence can be difficult to navigate for youth struggling with SEB difficulties. Without appropriate supports, these adolescents are at-risk of truancy, failing courses, being suspended or expelled, and dropping out (for a review see Mitchell, Kern, & Conroy, 2019). These detrimental academic outcomes serve as a call for prevention and intervention supports within secondary schools for at-risk youth. The identification of youth with academic, emotional, and behavioral difficulties is a necessary first step in Multi-Tiered

Corresponding author: Samantha M. Margherio, smkempker@gmail.com, 051 Porter Hall, 22 Richland Ave, Athens, OH, 45701.

Declaration of interest: none

Systems of Support (MTSS; Benner, Kutash, Nelson, & Fisher, 2013). Unfortunately, because MTSS universal screening practices have largely been developed and examined within elementary schools (Al Otaiba et al., 2009; VanDerHeyden, Witt, & Gilbertson, 2007), the feasibility and utility of tiered approaches to school-based intervention remains understudied in secondary school settings.

Investigations of common universal screening practices within secondary schools are necessary because many of the practices used at the elementary level may not translate well into middle and high schools (Fuchs, Fuchs, & Compton, 2010). Given the transient class schedules and established pattern of academic deficits present by middle and high school, universal screening practices such as teacher ratings and curriculum-based measures (CBM), though well-established in elementary settings, may be less feasible and useful as youth enter secondary school. Given this context, secondary school administrators report experiencing difficulties when making decisions regarding universal screening practices (Sansosti, Telzrow, & Noltemeyer, 2010). Thus, middle and high school mental health professionals and administrators are in need of developmentally-appropriate, psychometrically-sound, cost-effective universal screening tools to identify students at-risk of SEB and academic difficulties in order to connect them with services (Glover & Albers, 2007). Importantly, given the limited resources of several school districts, research delineating characteristics of youth identified by divergent screening methods is necessary to aid school administrators in developing efficient screening protocols to meet the district's needs. The current study seeks to provide information that can help guide secondary school administrators in their data-driven universal screening procedures by specifically contrasting three approaches.

Grade Point Average as an Indicator of Risk in Secondary Schools

When screening for students with SEB problems, school professionals commonly review students with academic impairment. In elementary school universal screenings, CBMs are recommended to identify youth in need of interventions (Benner et al., 2013). However, currently available CBMs have limited research support for use in secondary schools, and CBMs designed for secondary school use are difficult to create, implement, administer, and score (for a review, see Tindal, 2013), diminishing the feasibility of their use. Alternatively, Fuchs and colleagues (2010) recommend leveraging existing sources of data such as academic records within universal screening in middle and high schools because patterns of academic deficits may be well-established by adolescence. Indeed, current evidence suggests grade data are being used within universal screening practices in MTSS practices in secondary schools (Prewett et al., 2012). A descriptive report by the NHSC et al. (2010) outlined MTSS practices across 20 high schools and noted these schools used number of classes failed and poor academic test performance as indicators of need within the screening process.

Secondary students' grade point averages (GPA) are among the strongest predictors of high school graduation, even beyond attendance (Allensworth & Easton, 2007), indicating GPA may be an efficient tool for detecting youth at-risk of academic difficulties and/or school drop-out. Further, students with SEB difficulties often have lower GPAs than their peers

(Mcleod, Uemura, & Rohrman, 2013). However, a recent meta-analysis of universal screening procedures revealed only weak correlations between broadband screening tools and extant academic data across the gradespan (Allen, Kilgus, Burns, & Hodgson, 2019), although the analysis included few studies including self-report and secondary school samples. In sum, despite the potential promise of extant academic data as an efficient screening tool, its use may result in under-identification of students with SEB difficulties. Given the common use of grades to screen in secondary schools and questions about its utility, additional research exploring this hypothesis in secondary schools is needed. Alternatively, the integration of GPA data and broadband rating scales within a universal screening process may maximize the identification of heterogeneous risk among students.

Rating Scales as Indicators of Risk in Secondary Schools

Rating scales provide an alternative or supplement to grades as a screening tool. Several rating scales are available, although some have limitations that may preclude their utility at the secondary school level. First, rating scales with a cost per student, such as the BASC-3 Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2015), may be cost-prohibitive for many school districts (Mellard & Johnson, 2008). Second, ratings scales that only assess teacher-report, including the Systematic Screening for Behavior Disorders (SSBD 2nd Edition; Walker, Severson & Feil, 2014) and the Student Risk Screening Scale for Internalizing and Externalizing (SRSS-IE; Lane, Oakes, Cantwell, et al.,2016), miss important, incrementally valid information that may be obtained through self-report. Although in secondary schools the SSBD and SRSS-IE teacher-report have each been shown to identify youth at at-risk who would be missed by other rating scales or academic data (Erickson & Gresham, 2018; Lane, Oakes, Cantwell, et al., 2016), the lack of self-report data may result in under-identification of many at-risk students.

Complete reliance on teacher-report in secondary school screenings offers limitations related to limited awareness of students' difficulties and feasibility issues, and these limitations may be addressed by using self-report. Beginning in middle school, adolescents have several teachers who each have 100–150 students whom they see for a small portion of the week, leading to limited reliability and validity of teacher ratings (Evans, Allen, Moore, & Strauss, 2005). Secondary school teachers have little supervision of students outside of their classroom, and thus may not be accurate reporters of many of adolescents' SEB difficulties. Further, it would be a time-consuming and costly endeavor for secondary teachers to provide ratings for all, or even half, of their students. Alternatively, self-report indices may offer a more feasible and valid approach to screening. Although adolescent self-report of externalizing behaviors may have limited validity (Smith et al., 2007), students may be the optimal source for gathering information related to internalizing problems (Aebi et al., 2017; Silverman & Ollendick, 2005).

One universal screening measure that is free and includes both a teacher and self-report version is the SDQ (Goodman, Meltzer, & Bailey, 1998). The SDQ is a brief measure of internalizing and externalizing difficulties and prosocial skills. The teacher version of the SDQ has demonstrated convergent validity with another teacher-rated screening measure within a middle school sample (i.e., SRSS-IE; Lane, Oakes, Carter, Lambert, & Jenkins,

2013); however, the time-consuming nature of the SDQ teacher version (e.g., 25 items per student) is a limitation of the measure (Lane, Kalberg, Parks & Carter, 2008). The reliability and construct validity of the self-report version of the SDQ have been confirmed within secondary schools across several European countries (see Ortuño-Sierra et al., 2015). Convergent validity of the self-report SDQ with other indices of mental health, diagnostic status, and teacher-report have been supported in multiple clinical and non-clinical samples across Europe (Aebi et al., 2017; Goodman, Ford, Simmons, Gatward, & Meltzer, 2000).

Despite the assertion that "there is extensive inquiry across the grade span" regarding the SDQ as a screening measure (Lane, Oakes, Lusk, Cantwell, & Schatschneider, 2016, p. 160), research has yet to clarify the characteristics of youth identified as at-risk by the self-report SDQ whom would otherwise be missed by other measures. Using the SDQ in a European sample, Goodman and colleagues (2000) determined ratings from teachers were more sensitive to identifying youth with any disorder than parent or self-ratings, with a particular advantage for identifying youth with externalizing disorders. They also reported that adding adolescent self-report to teacher-report resulted in increased sensitivity to identify youth with a depression or anxiety diagnosis, but the sensitivity for identifying externalizing disorders remained unchanged. However, because teacher-report alone and self-report alone demonstrated comparable sensitivity to internalizing difficulties, Goodman and colleagues concluded self-report was minimally useful in screening practices, even for identification of internalizing problems. These findings are limited because ICD-10 diagnoses were determined by gathering data from the same three sources as the ratings and the weight placed on data from each source when determining the diagnoses may have influenced the sensitivity analyses. This may at least partially explain why their conclusion about self-report of internalizing problems contrasts with that of other researchers (Aebi et al., 2017; Silverman & Ollendick, 2005). Additionally, teachers may only detect severe internalizing and externalizing difficulties (Splett et al., 2019) limiting their use identifying students with problems, but without diagnoses. Finally, adolescents without clinical diagnoses tend to report more SEB problems than observer informants (Achenbach, Dumenci, & Rescorla, 2002; Vierhaus & Lohaus, 2008). Examining the utility of the self-report SDQ for identification of problems in various domains, including sub-diagnostic threshold problems, may lead to different conclusions that are more aligned with school-based screening than the identification of students with a diagnosis. Thus, research investigating both the teacher and student versions of the SDQ, as well as GPA are needed to inform the choices administrators face when selecting screening tools, particularly at the understudied secondary level.

Further, research investigating the SDQ self-report in the U.S. is lacking. One study in the U.S. demonstrated self-report on the SDQ produced similar diagnostic accuracy as a structured clinical interview among in-patient adolescents (Kovacs & Sharp, 2014). To our knowledge, the only other investigation of the SDQ self-report in the U.S. demonstrated support for an internalizing, externalizing, and prosocial factor solution (Ruchkin, Jones, Vermeiren, & Schwab-Stone, 2008); however, no study has investigated the SDQ self-report as a school-based screening tool within the U.S. Given the potential cross-cultural variations in SDQ self-ratings (Ravens-Sieberer, Erhart, Gosch, & Wille, 2008), research is needed to explore the utility of the SDQ self-report within screening procedures in the U.S.

Screening Cutoff Scores

Within the MTSS model, universal screening practices often identify students with “some risk” and “high-risk” (Benner et al., 2013), and these multi-level approaches to risk can help connect students to appropriate tiers of intervention. Many rating scales include multiple cutoffs, although the implications of using them has received little empirical examination. High-risk cutoffs on teacher ratings identified a larger portion of middle than high school students (Lane, Oakes, Cantwell, et al.,2016), whereas the opposite has been found for self-report (Dever, Kamphaus, Dowdy, Raines, & DiStefano, 2013). Further, the utility of different cutoffs on rating scales remains unclear, as the moderate- and high-risk thresholds on the teacher-rated SRSS (Lane et al., 2008) and self-report SDQ (Renshaw, 2019) have functioned similarly in identifying students with academic difficulties. However, research investigating the utility of different cutoffs for GPA has found dramatic decreases in likelihood of graduation if a student’s GPA falls below a 2.0 or 1.0, with likelihood of high school graduation dropping to 75% and 25%, respectively (Allensworth & Easton, 2007). Thus, it is important to systematically study the implications of using cutoffs for moderate- or high-risk within universal screening practices in secondary schools.

Demographic Characteristics and Screening

There is some evidence that teacher ratings, but not student ratings, are unduly influenced by demographic factors such as gender and socio-economic status (SES; Dever, Raines, Dowdy, & Hostutler, 2016; Sargisson, Stanley, & Hayward, 2016). This pattern is reflected within special education placements, although student risk status based on self-report does not demonstrate such elevated disproportions by gender or SES (Dever et al., 2016). However, some studies have reported null findings regarding the influence of demographic variables on ratings (for a review see De Los Reyes & Kazdin, 2005). Self-report screeners provide an alternative source to teacher report that may mitigate overrepresentation of males and students of low SES in school-based services. Examinations comparing the rates of identification by source are needed to address this finding.

Summary and Purpose

Although we have learned a great deal about screening for students with SEB at the elementary level, many questions remain for screening in middle and high schools. In order to inform cost-efficient, developmentally-appropriate universal screening practices, research is needed that examines the relative value of various sources of screening data. Specifically, given the context of secondary schools, coupled with contrasting claims in the literature regarding the utility of self-report within screening procedures, questions exist regarding the utility of teacher-report compared to student self-report for the identification of SEB problems. In contrast to gathering ratings, archival data offers a cheap and convenient source of screening data and has led to administrators relying on grade data for screening despite findings indicating limited value for identifying students with SEB problems. Although reports indicate grades may not be an effective source for identifying these students, their utility has not been compared to teacher and self-ratings. All three sources of screening data need to be considered in relation to each other and in combination to determine which

students are identified and missed with each source and combination of sources. These findings can inform the selection of screening tools and may enlighten methods for considering data collected from multi-informant, multi-method screening procedures. This research also needs to take into account that the value of sources may vary as a function of the types of problems being considered. Further, current literature includes mixed findings in relation to the use of available cutoff scores for the determination of risk for teacher-ratings, self-ratings, and grades. Research is needed to demonstrate the impact on rates of identification within universal screening when using varying cutoffs.

In this study we examined the characteristics of adolescents identified by self- and teacher-report using the SDQ (Goodman et al., 1998) and GPA at moderate and high levels of risk. We answer the following questions: (1) What is the association and agreement across GPA and teacher- and student-report? (2) How does identification of at-risk students differ when using moderate or high severity cutoff scores across measures? (3) What are the characteristics of adolescents identified as at-risk by GPA, teacher and self-report measures? We predicted students rather than teachers would report more internalizing and externalizing difficulties, and therefore more students would be identified as at-risk by student-report than by teacher-report (Achenbach et al., 2002; Vierhaus & Lohaus, 2008).

Method

Participants

The school district targeted for recruitment was located within a metropolitan area of Southeast Ohio. A total of 2,350 students across two middle school (grades 6 through 8) and one high school (grades 9 through 12) were included. The sample included 49.5% middle school students, 46.5% females, and most youth identified as White (94.3%). The sample was socioeconomically diverse, and 42.6% of students were eligible for free and reduced lunch (FRL).

Procedures

Data were collected in the context of an epidemiologic study, the Project to Learn about Youth-Mental Health (masked, 2014). All procedures were approved by the university research board and school administrators. Letters describing the study and the passive consent process were mailed to all guardians of students. Families could decline participation by contacting the project coordinator or principal. The passive consent technique was used in order to attain a sample large enough to be representative of the participating schools and larger district.

Teachers were provided a list of student names for whom to complete the surveys and offered $5 per student. All teachers completed the SDQ digitally using the Research Electronic Data Capture (REDCap) system (Harris et al., 2009). Investigators used the following guidelines to select teacher informants: (1) each teacher was assigned no more than 30 students; (2) teachers of study halls, physical education, music, or on-line courses were not included; (3) the teacher of a student's first period was invited first; if that teacher declined or the student was not on campus the first period, the teacher of the second period

class was invited, and this process continued until there was a teacher match for each student.

The student screening included the self-report version of the SDQ obtained in group format using paper-and-pencil versions. Middle school students completed the measures during the first period and high school students completed the measures during a morning meeting on a late start day. If students were not present, school counselors attempted to have them complete the survey at another time. Ratings were obtained during the 2015–2016 school year, with ratings for high school students (teacher and student) obtained in the spring, and ratings for middle school students obtained in the fall (teacher) and spring (student).

Missing data.—The dataset included 1,161 middle and 1,189 high school students for whom demographic and screening data are available, representing 85% and 69% of the middle and high school student bodies, respectively. Teacher ratings were missing if teachers declined to participate, the student was not attending on-campus, or the student withdrew before the screening date. Student ratings were missing if parents actively declined participation (6 in middle school; 28 high school), the student was absent on the day of screening, or the student attended but actively declined to complete the surveys (we do not have a way to distinguish the latter two groups). Of the 34 students whose parents actively declined participation (non-participants), 41% were female, 91% were White, and 41% qualified for FRL. Non-participating students were more likely to be in middle than high school, $\chi^2(1, N = 3,103) = 9.88$, $p < .01$. There were no significant differences in gender, race, or FRL status between participants and non-participants.

Among participants, data were not missing at random. Missing student-report ($n = 569$) was more common than missing teacher-report ($n = 175$). Of students for whom teacher-report was available, 29% of students rated as at-risk by the teachers did not have self-report data; in contrast, 15% of students rated as low-risk by the teachers did not have self-report data, $\chi^2(1, N = 2,928) = 60.23$, $p < .01$. Additionally, 25% of students meeting at-risk criteria due to their GPA did not have student-report, whereas 13% of students with low-risk GPAs did not have student-report, $\chi^2(1, N = 2,942) = 65.78$, $p < .01$. Finally, 7% of middle school students did not complete self-reports whereas 26% of high school students did not complete self-reports, $\chi^2(1, N = 3,069) = 176.58$, $p < .01$. There were no significant differences between students with or without teacher-report.

Measures

GPA.—Unweighted GPAs were collected for every student during each quarter of the school year during the study. Second quarter grades for math, science, English/Language arts, and social studies were used to determine the student's GPA for this study because this time point was closest to the times we collected student- and teacher-reports. Prior research has found dramatic decreases in likelihood of graduation if a students' GPA falls below a 2.0 or 1.0 (Allensworth & Easton, 2007). Thus, on the 4.0 GPA scale, we defined GPAs below 2.0 or 1.0 as moderate- or high-risk, respectively.

SDQ.—The SDQ is a 25-item mental health screening measure for children and adolescents (Goodman et al., 1998) with five subscales (i.e., emotional symptoms, conduct problems,

hyperactivity-inattention, peer relations, and prosocial behavior) of five questions each. Each item is rated on a three-point scale: *not true (0), somewhat true (1),* or *certainly true (2)*. The first four scales are summed to form a total difficulties score ranging from 0 to 40. Consistent with recommendations from Goodman, Lamping, and Ploubidis (2010), teachers' and students' SDQ conduct problems and hyperactive-inattention subscales were summed to create an externalizing variable, and peer relationships and emotional problems were summed to create an internalizing variable. Prosocial behaviors were examined separately. These composite variables and prosocial behaviors were examined dimensionally within the primary analyses. Secondary analyses examining rates of categorical risk status by specific problems included data from the original subscales. Given the limitations of Cronbach's alphas as indices of internal consistency, including the strict tau equivalence assumption and potential inaccuracies of the point estimate (see Dunn, Baguley, & Brunsden, 2014), Cronbach's alphas and their 95% confidence intervals as well as omega values are reported as indices of internal consistency. In the current sample, internal consistency estimates for the teacher version were: α = .79 (.78, .80), ω = .80 for internalizing problems, α = .87 (.86, .88), ω = .87 for externalizing problems, α = .87 (.86, .88), ω = .87 for prosocial behaviors, and α = .86 (.86, .87), ω = .87 for the total difficulties. For the student version, estimates of internal consistency were: α = .74 (.72, .75), ω = .74 for internalizing problems, α = .78 (.76, .79), ω = .77 for externalizing problems, α = .69 (.67, .71), ω = .70 for prosocial behaviors, and α = .81 (.80, .82), ω = .81 for the total difficulties.

Risk status was defined using pre-established cutoffs of the SDQ total difficulties score (see Goodman & Goodman, 2009). The moderate-risk and high-risk cutoffs are meant to capture the most extreme 20% and 10% of students in the normative population, respectively. Using these cutoffs, moderate-risk was defined as a score of 12 or higher on teacher-report and 16 or higher on self-report, and high-risk was defined as a score of 16 or higher on teacher-report and 20 or higher on self-report. Thus, "moderate-risk" in this study is indicative of adolescents rated as having moderate or greater risk. Next, we examined the portion of youth with high-risk subscale scores captured by each screening measure. Moderate- and high-risk cutoffs for the four subscales making up the internalizing and externalizing dimensions were used in the secondary analyses.

Defining risk groups.—In this multi-method, multi-informant universal screening study, students could be identified as at-risk by just one screening measure, or by multiple screening measures. In order to compare the differences among youth identified by one, or combinations of the screening measures, two variables were created and are represented visually, and proportionately in Figures 1 and 2. Students were categorized into one of eight risk groups; low-risk (0), or at-risk by GPA only (1), teacher-report only (2), student-report only (3), GPA and teacher-report (4), GPA and student-report (5), teacher- and student-report (6), or all three screening measures (7). Students were categorized into a risk group if they exceeded the moderate-risk (Figure 1) or high-risk (Figure 2) cutoff.

Data Analytic Plan

Aim 1: Association between measures.—Zero-order correlations were calculated to assess the association between GPA and teacher and student SDQ total difficulties scores.

Although GPA and teacher-report demonstrated moderate skew, skewness coefficients were within acceptable limits (e.g., less than +/− 1.20; Gravetter & Wallnau, 2014). Kappa coefficients were used to examine cross-informant agreement between student-report, teacher-report, and GPA using the moderate-risk cutoff.

Aim 2: Effect of changing level of risk.—To examine how the identification of at-risk students differ when using moderate versus high severity cutoff scores, the number of students identified by each variable using each criterion were calculated. The agreement analyses described for Aim 1 were repeated using the high-risk cutoff.

Aim 3: Characteristics and percent of students identified by different measures.—For Aim 3, the sorted groups (see Figures 1 and 2 for a proportional representation) were contrasted on demographic characteristics, teacher-ratings, self-ratings, and grades. First, a series of Pearson chi-squared analyses were conducted to compare demographic variables (i.e., gender, school level, FRL status) across the risk groups. Second, multivariate analysis of variances (MANOVAs) were conducted to assess differences across students identified by various screening measures. Teacher- and student-rated prosocial behaviors and internalizing and externalizing problems were entered as outcome variables. Prior to conducting contrasts and multivariate analyses, youth identified by GPA, teachers, and students combined were removed from the high-risk group because there were only six individuals in this group, which does not allow for a sufficient number of observations per cell in the MANOVA (Huberty & Petoskey, 2000). Although the assumption of multivariate normality was violated, having a large sample makes the MANOVA robust to this violation (Huberty & Petoskey, 2000). Box's M revealed unequal cell sample sizes; thus, Pillai's Trace, was used to assess for statistical significance of the model (Morrison, 2005). A post-hoc power analysis revealed the MANOVA was sufficiently powered to detect moderate to large differences.

Next, a series of contrasts were conducted in order to compare SEB difficulties in adolescents identified as at-risk by various screening measures, as well as between adolescents identified as at-risk and those identified as not at-risk. Hedges' g effect sizes were calculated for each contrast as it takes into account different sample sizes (Hedges, 1981). The Bonferroni correction was used to adjust the alpha level for 12 contrasts across six variables (α =.0007). Finally, in order to demonstrate the portion of youth with difficulties identified by each screening measure, we assessed the percentage of youth meeting risk status for academic, emotional problems, peer problems, conduct problems, and hyperactivity-inattention difficulties captured by GPA, teacher-report, and student-report.

Results

Aim 1: Association Between Measures

Table 1 displays the correlations and indices of agreement between measures. GPA had a moderate negative relationship with teacher-report, and a weak negative relationship with student-report. Teacher- and student-report had a weak positive relationship. Per McHugh's (2012) kappa interpretation standards, there was minimal agreement among GPA and

teacher risk status, but no agreement between GPA and student risk status and teacher and student risk status.

Aim 2: Effect of Changing Level of Risk

Using the moderate criteria, 46% of the sample was identified by one or more methods, with GPA risk status identifying the most adolescents at moderate-risk (27%), self-report identifying 24% of the sample as at-risk, and teacher-report identifying 15% of the sample as at-risk. Increasing the cutoff reduced the portion of youth identified as at-risk from 46% to 20%, with self-report identifying the largest portion as high-risk (10%) and teacher and GPA identifying 7% each. Increasing the cutoff resulted in the smallest change in number of youth identified by teacher-report (15% to 7%) relative to GPA or student-report. Additionally, the number of students *uniquely* identified as at-risk by teacher-report only dropped from 96 to 93 students, indicating if students were uniquely identified by teachers, teachers were likely rating them as high-risk.

Increasing the risk cutoff reduced agreement among indices (see Table 1). The decreased agreement was also evident by proportionately fewer youth being identified by multiple screening measures. Because findings from the analyses using the high-risk cutoff generally trended in the same direction as those from the moderate groups, and the high-risk cutoff identified a more feasible portion of students in need of more extensive evaluation, only results from the high-risk group are presented below.

Aim 3: Characteristics and Percent of Students Identified by Each Measure

Demographic characteristics for the moderate- and high-risk groups are represented in Figures 1 and 2, respectively. The chi-square tests revealed significant associations between high-risk status and gender ($\chi^2(7) = 58.45$, $p < .01$), school level ($\chi^2(7) = 26.67$, $p < .01$), and FRL status ($\chi^2(7) = 67.24$, $p < .01$). Among high-risk groups, self-report identified the most females as at-risk whereas teacher-report identified the fewest females. GPA identified the most high school students as at-risk whereas self-report identified the most middle school students. Students qualifying for FRL were much more likely to be identified as at-risk than low-risk by any screening measure, and were more likely to be identified as at-risk by multiple screening measures than by any one screening tool. Chi-squared analyses were not conducted to examine differences in race across groups due to low cell frequencies; however, descriptive data revealed the highest proportion of minority youth were identified by teachers only (10%), followed by GPA only (8%), and teacher and GPA (7%). No minority youth were identified as high-risk by both GPA and self-report, and 5% of low risk youth were racial minorities.

Table 2 displays the means of each risk groups' internalizing and externalizing ratings and prosocial behaviors. The MANOVA revealed significant differences in teacher- and student-rated externalizing and internalizing difficulties and prosocial behaviors across high-risk groups, ($F(36, 14022) = 66.35$, $p < .01$; Pillai's Trace = .87). Table 3 displays the Hedges' g effect sizes for the series of contrasts conducted in order to understand differences in characteristics of students identified as at-risk by GPA, teacher-report, and student-report, and differences between these students and low-risk students.

The contrasts revealed GPA identified students with fewer teacher- and student-rated externalizing problems, and fewer student-rated internalizing problems than students identified by teacher- or self-report. However, students identified as at-risk by GPA did not differ from other at-risk students in teacher-rated internalizing problems or prosocial behaviors. Students identified as at-risk by teacher-report had fewer student-rated internalizing difficulties than other at-risk students, but they did not differ from other at-risk groups on student-rated externalizing difficulties or prosocial behaviors. Thus, teacher-report performed similarly to other measures in identifying youth with student-rated externalizing difficulties. Student-report uniquely identified students with the greatest self-reported internalizing difficulties, indicating student-report alone captures youth reporting the most severe internalizing problems. GPA, teacher-report, and student-report each identified students with greater internalizing and externalizing difficulties and fewer prosocial behaviors than students in the low-risk group, indicating each screening measure differentiated between low- and at-risk students.

Table 4 represents the proportion of students whose subscale scores met the subscale risk threshold, who were captured as at-risk by the broader GPA and SDQ total difficulties cutoffs. GPA identified a small portion of students with teacher-rated emotional (11%) and peer problems (14%), and even fewer students with student-rated emotional (5%) and peer problems (5%). GPA identified slightly larger portions of students with teacher-rated conduct (20%) and hyperactivity problems (19%), or with student-rated conduct (13%) and hyperactivity problems (11%). Using the total difficulties cutoff, teacher-report identified most students with teacher-reported conduct problems above the subscale cutoff (65%); however, teacher-report identified only 37% of students with teacher-rated hyperactivity problems, and 50% and 40% of students with teacher-rated emotional and peer problems, respectively. Taken together, if teachers identified a student as at-risk, it was likely due to concerns of conduct problems rather than hyperactivity, peer, or emotional problems. Teacher-report did not identify most students with student-rated difficulties, and specifically did not identify 93% of students with student-rated emotional problems. Teachers were more likely to rate males than females as at-risk on peer, hyperactivity, and conduct problems. Student report identified a greater proportion of students with teacher-rated difficulties than the converse. Student-report captured nearly half of students with self-reported conduct (57%), hyperactivity (44%), emotional (44%), and peer problems (50%). Females were more likely than males to rate themselves as at-risk in internalizing difficulties.

Discussion

This study advances the science of universal screening practices in secondary schools by comparing three common secondary school screening approaches: GPA, teacher-report, and student-report. The results provide researchers and school administrators with information about the students identified by each approach and the feasibility of moderate- and high-risk cutoff criteria. Further, this study was the first to examine the SDQ, a free and brief screening tool, as a multi-informant universal screening approach in secondary schools within the U.S.

Our study offered three main findings with important implications for universal screening in secondary schools. First, we found that using a moderate-risk cutoff within multi-informant, multi-method universal screening procedures likely identifies more students than a school district can feasibly address. Almost half of the sample (46%) was identified as at-risk by at least one screening tool when using the moderate-risk cutoff. Potentially unsuitable cutoffs and modest agreement among measures likely contributed to such a large proportion of youth being identified as at-risk. The SDQ cutoffs are based on a large European sample and are meant to capture the most severe 20% (moderate) and 10% (high) at-risk youth (see Goodman & Goodman, 2009). However, in our large, socioeconomically diverse U.S. sample, the moderaterisk criteria captured 15% of students on teacher-report and 24% of students on self-report, and the high-risk criteria captured 7% and 10% of youth, respectively. Similar variations from the normative percentages have been found in other countries (see Ravens-Sieberer et al., 2008). This pattern of findings suggests the SDQ may require additional research within the U.S. to determine cutoffs that resemble intended portions of the population. Due to the modest agreement among measures, many students were identified by just one screening tool, resulting in a large portion of youth identified as at-risk. This finding is consistent with other investigations of screening practices, which have found as many as 48% of high school students in accelerated courses were identified as at-risk by either academic performance, self-report, or teacher nomination (Suldo et al., 2019). Similarly, in a large sample of elementary and middle school students, the limited overlap between teacher-report and academic data resulted in 38% of students being identified by only one measure of risk (Miller et al., 2015). Therefore, this moderate cutoff and use of multiple measures is unlikely to be a feasible approach to screening for at-risk students. Indeed, in one study involving screening for suicidality in high schools, school administrators chose to discontinue screening efforts when 29% of the student body screened positive for suicidality, stating the screening process “is totally unrealistic given the numbers we are getting” (Hallfors et al., 2006, p. 284). Our findings suggest using high-risk criteria may be more useful than moderate-risk criteria within universal screening practices.

Second, student report identified the largest number of youth as at-risk, and their report was more sensitive to capturing students with teacher-rated concerns than the converse. Student self-report was uniquely valuable in identifying students reporting elevated internalizing difficulties. Although prior researchers have offered contrasting claims regarding the sensitivity of self-report for identification of youth with mental health diagnoses (e.g., Aebi et al., 2017; Goodman et al., 2000), sensitivity to youth with diagnoses is less relevant to universal screening than clinical evaluations. This difference in methods may account for differences between our findings and those of Goodman and colleagues (2000) who reported little value in obtaining selfreport ratings. Our findings suggest adolescent self-report may be more cost-effective than teacher-report for identifying youth with either self- or teacher-rated concerns in both internalizing and externalizing domains.

Third, students identified as at-risk by GPA had greater emotional and behavioral difficulties than the low-risk group, yet these students had significantly fewer externalizing problems by teacher and self-report and fewer internalizing problems by student-report than students identified as at-risk by one or both of the raters. Collectively, these findings support prior evidence that students with poor GPAs may have elevated emotional and behavioral

difficulties relative to their peers (e.g., McLeod et al., 2013), yet GPA is only minimally associated with broadband screening tools (Allen et al., 2019). Our findings indicate little value in using GPA for screening for emotional and behavioral problems, despite its current use in screening. This finding is unfortunate as GPA is the cheapest and easiest data to obtain. Still, there may be value in including GPA in a screener in order to identify students with SEB problems and academic failure. Fifty-six students were identified as at high-risk based on GPA and either self- and/or teacher-report. A strong case could be made that these 56 students out of the 2,350 screened warrant the most immediate attention.

Additionally, we found male and low SES students were overall more likely to be identified as at-risk than their peers, consistent with previous findings that these youth receive higher ratings on screening measures and are more likely to receive school-based services (Dever et al., 2016; Sargisson et al., 2016). However, our findings suggest that integration of self-report in universal screenings may mitigate the gender disproportion present in educational referral rates. Student self-report data indicated that 57% of those at high risk were female and teacher-report indicated that only 35% of those at high risk were female. Thus, we replicated the gender difference in teacher ratings and found that within the same sample, the gender difference is not reflected in student ratings. Part of this difference may be attributable to teachers' concerns being largely driven by conduct problems. Interestingly, youth of low SES backgrounds were more likely to be identified as at-risk by multiple screening measures, suggesting these youth experience difficulties that inhibit academic success, and are noticeable to both their teachers and themselves.

Our results indicate high school students were more likely than middle school students to be identified as at-risk. These findings contrast with research evaluating moderate- and high-risk criteria on the teacher-rated SRSS-IE (Lane, Oakes, Cantwell, et al., 2016). Questions on the SRSS-IE and the SDQ may differ in a manner that are considered differently by raters across grades. Further, the hypothesis that high-risk adolescents likely drop out and therefore are not as represented in high schools was not supported in our sample (Lane, Oakes, Cantwell, et al., 2016). These contrasting results speak to the need for research examining the transition of risk from middle to high school.

Implications

In spite of the benefits of using all three sources in secondary schools, there may be limitations regarding the resources available to collect and analyze this information. As a result, it is important to consider the degree of loss of information by eliminating any one of the sources. The percentage missed would be greatest if self-report were eliminated (37%) and less if GPA were eliminated (24%) and even less if teacher-report were eliminated (20%). As a result, if there is a need to eliminate one source, the elimination of teacher-report would result in the fewest number of students missed compared to when all three measures are used.

If the goal of screening is to capture youth with internalizing difficulties, teacher-report may be of minimal utility, as only a small portion of youth were rated by teachers as having internalizing difficulties, and teacher-report missed the large majority of youth who self-reported internalizing difficulties. If the goal of screening is to capture youth with

externalizing difficulties, teacher ratings may be useful, in conjunction with self-report. However, teacher ratings disproportionally identified students with conduct problems, indicating students identified by teacher-report may already be identified as at-risk by teacher nominations or office disciplinary referrals. Given that teachers have cited the length of the SDQ (e.g., 25 items) as a barrier to its use (Lane et al., 2008), school administrators may consider streamlined methods for obtaining teacher-report, such as through a nomination procedure. Although in elementary universal screening procedures, teacher ratings enhance identification of students beyond teacher-nomination (Eklund et al., 2009), teacher-nomination within secondary schools has been recommended in place of rating scales (Fuchs et al., 2010), and our results speak to the need for more research comparing these procedures in secondary schools.

Finally, if administrators want an approach that efficiently identifies students with any type of SEB problem, then the most efficient approach appears to be to self-report with high-risk criterion. Self-report identifies the most at-risk students, captures students with the most severe self-reported internalizing difficulties, and is more sensitive to teacher-rated difficulties than teacher-report is to student-rated difficulties. As described earlier, once the at-risk group is identified by self-report, administrators may wish to review that sample with GPA data and prioritize those with both SEB problems and failing grades.

Limitations and Future Directions

Limitations of this study include missing data patterns and reliance on one school district for data. Self-report was missing from adolescents more likely to be at-risk on teacher-report and GPA than students who completed the ratings. Although this is a limitation of the representativeness of our sample to all secondary school students, this lack of self-report may be representative of the information available when conducting universal screening in secondary schools as our reasons for missing data are likely to similarly impact screening practices. Additionally, the lack of racial and ethnic diversity in our sample may limit the generalizability of our findings. However, our study was the first step in examining the SDQ self-report as a screener in the U.S., and future studies may build upon these findings with racially and ethnically diverse samples, in addition to continuing to examine the SDQ with socioeconomically diverse samples. Middle school teachers and students completed the screeners at different times of the year, which may have affected the level of agreement between student and teacher-report due to changes in the student’s needs and difficulties over the course of the year. Further research of cutoffs may be helpful, especially for GPA as normative GPA is likely to vary substantially between schools. Additional data related to office disciplinary referrals, attendance, or other mental health and education information would have increased confidence in and the breadth of conclusions regarding these findings. Future research is warranted evaluating how the screening measures used in this study relate to other measures of risk gathered over time, such as attendance, dropout rates, and future academic success.

Conclusion

Universal screening for academic, emotional, and behavioral difficulties in secondary schools is a necessary first step in identification of at-risk youth. The current study sought to inform universal screening practices by comparing multiple screening tools within a large, socioeconomically-diverse sample. Our results demonstrate that risk identification varied substantially across screening options, and use of a moderate-risk cutoff resulted in nearly half of all students being identified as at-risk. Integration of multiple screening methods with a high-risk cutoff resulted in a more feasible number of students to be further addressed. Further, our results demonstrate that reliance on teacher-report and extant academic data may allow adolescents experiencing internalizing and externalizing difficulties to fall through the cracks. Alternatively, among screening tools investigated, student rating scales may offer the most efficient means for identifying youth with social, emotional, and behavioral difficulties; thus, including self-report in universal screening practices in secondary schools is an important practice to identify at-risk youth.

Acknowledgement:

Data were obtained in a project supported by the Disability Research and Dissemination Center (DRDC) through its Grant Number 5U01DD001007-03, FAIN No. U01DD001007 from the Centers for Disease Control and Prevention (CDC). The findings and conclusion of the article are solely the responsibility of the authors and do not necessarily represent the official views of the DRDC or the CDC.

References

Achenbach T, Dumenci L, & Rescorla L (2002). Ten-year comparisons of problems and competencies for national samples of youth: Self, parent, and teacher reports. Journal of Emotional and Behavioral Disorders, 10 194–203. doi:10.1177/10634266020100040101

Aebi M, Kuhn C, Banaschewski T, Grimmer Y, Poustka L, Steinhausen H, & Goodman R (2017). The contribution of parent and youth information to identify mental health disorders or problems in adolescents. Child and Adolescent Psychiatry and Mental Health, 11, 11–23. doi:10.1186/s13034-017-0160-9 [PubMed: 28286549]

Al Otaiba S, Connor C, Foorman B, Schatschneider C, Greulich L, & Sidler J (2009). Identifying and intervening with beginning readers who are at-risk for dyslexia: Advances in individualized classroom instruction. Perspectives on Language and Literacy, 35, 13–19. [PubMed: 25598861]

Allen A, Kilgus S, Burns M, & Hodgson C (2019). Surveillance of internalizing behaviors: A reliability and validity generalization study of universal screening evidence. School Mental Health, 1–1, 194–209. doi:10.1007/s12310-018-9290-3

Allensworth E, & Easton J (2007). What matters for staying on-track and graduating in Chicago public high schools: A close look at course grades, failures, and attendance in the freshman year. Research Report. Consortium on Chicago School Research. doi: 10.1080/10824669.2013.745181

Benner G, Kutash K, Nelson J, & Fisher M (2013). Closing the achievement gap of youth with emotional and behavioral disorders through multi-tiered systems of support. Education and Treatment of Children, 36, 15–29. doi:10.1353/etc.2013.0018

De Los Reyes A, & Kazdin A (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. Psychological Bulletin, 131, 483–509. doi:10.1037/0033-2909.131.4.483 [PubMed: 16060799]

Dever B, Kamphaus R, Dowdy E, Raines T, & DiStefano C (2013). Surveillance of middle and high school mental health risk by student self-report screener. Western Journal of Emergency Medicine, 14, 384–391. [PubMed: 23997848]

Dever B, Raines T, Dowdy E, & Hostutler C (2016). Addressing disproportionality in special education using a universal screening approach. Journal of Negro Education, 85, 59–71. doi: 10.7709/jnegroeducation.85.1.0059

Dunn TJ, Baguley T, & Brunsden V (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105, 399–412. doi: 10.1111/bjop.12046 [PubMed: 24844115]

Eklund K, Renshaw T, Dowdy E, Jimerson S, Hart S, Jones C, & Earhart J (2009). Early identification of behavioral and emotional problems in youth: Universal screening versus teacher-referral identification. The California School Psychologist, 14, 89–95.

Erickson M, & Gresham F (2018). Measuring teachers' perceptions of student behavior using the Systematic Screening for Behavior Disorders in middle school students. Journal of Emotional and Behavioral Disorders, 27, 119–128. doi:10.1177%2F1063426618763110

Evans SW, Allen J, Moore S, & Strauss V (2005). Measuring symptoms and functioning of youth with ADHD in middle schools. Journal of Abnormal Child Psychology, 33, 695–706. doi:10.1007/s10802-005-7648-0 [PubMed: 16328745]

Fuchs L, Fuchs D, & Compton D (2010). Rethinking response to intervention at middle and high school. School Psychology Review, 39, 22–28.

Glover T, & Albers C (2007). Considerations for evaluating universal screening assessments. Journal of School Psychology, 45, 117–135. doi:10.1016/j.jsp.2006.05.005

Goodman A, & Goodman R (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. Journal of the American Academy of Child & Adolescent Psychiatry, 48, 400–403. doi:10.1097/CHI.0b013e3181985068 [PubMed: 19242383]

Goodman A, Lamping D, & Ploubidis G (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. Journal of High Child Psychology, 38, 1179–1191. doi:10.1007/s10802-010-9434-x

Goodman R, Ford T, Simmons H, Gatward R, & Meltzer H (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. British Journal of Psychiatry, 177, 534–539. doi:10.1192/bjp.177.6.534 [PubMed: 11102329]

Goodman R, Meltzer H, & Bailey V (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. European Child & Adolescent Psychiatry, 7, 125–130. doi:10.1007/s007870050057 [PubMed: 9826298]

Gravetter FJ, & Wallnau LB (2005). Essentials of statistics for the behavioral sciences. Belmont. CA: Wadsworth.

Hallfors D, Brodish P, Khatapoush S, Sanchez V, Cho H, & Steckler A (2006). Feasibility of screening adolescents for suicide risk in "real-world" high school settings. American Journal of Public Health, 96, 282–287. doi:10.2105/AJPH.2004.057281 [PubMed: 16380568]

Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, & Conde J (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics, 42, 377–381. doi: 10.1016/j.jbi.2008.08.010 [PubMed: 18929686]

Hedges L (1981). Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics, 6, 107–128. doi:10.2307/1164588

Huberty C & Petoskey M (2000). Multivariate analysis of variance and covariance In Tinsley HE & Brown SD (Eds), Handbook of applied multivariate statistics and mathematical modeling (pp. 183–208). Cambridge, MA: Academic Press. doi:10.1016/B978-012691360-6/50008-2

Kamphaus R & Reynolds C (2015). BASC-3 behavioral and emotional screening system manual. Circle Pines, MN: Pearson.

Kovacs S, & Sharp C (2014). Criterion validity of the Strengths and Difficulties Questionnaire (SDQ) with inpatient adolescents. Psychiatry Research, 219, 651–657. doi:10.1016/j.psychres.2014.06.019 [PubMed: 25048754]

Lane K, Kalberg J, Parks R & Carter E (2008). Student Risk Screening Scale: Initial evidence for score reliability and validity at the high school level. Journal of Emotional and Behavioral Disorders, 16, 178–190. doi:10.1177/1063426608314218

Lane K, Oakes W, Cantwell E, Schatschneider C, Menzies H, Crittenden M & Messenger M (2016). Student Risk Screening Scale for internalizing and externalizing behaviors: Preliminary cut scores to support data-informed decision making in middle and high schools. Behavioral Disorders, 42, 271–284. doi:10.1177/1063426615618624

Lane K, Oakes W, Carter E, Lambert W, & Jenkins A (2013). Initial evidence for the reliability and validity of the Student Risk Screening Scale for Internalizing and Externalizing Behaviors at the middle school level. Assessment for Effective Intervention, 39, 24–38. doi: 10.1177/1534508413489336

Lane K, Oakes W, Lusk M, Cantwell E, & Schatschneider C (2016). Screening for intensive intervention needs in secondary schools: Directions for the future. Journal of Emotional and Behavioral Disorders, 24, 159–172. doi:10.1177/1063426615618624

McHugh M (2012). Interrater reliability: the kappa statistic. BiochemiaMedica, 22, 276–282.

McLeod J, Uemura R, & Rohrman S (2012). Adolescent mental health, behavior problems, and academic achievement. Journal of Health and Social Behavior, 53, 482–497. doi: 10.1177/0022146512462888 [PubMed: 23197485]

Mellard D, & Johnson E (Eds.). (2008). RTI: A practitioner's guide to implementing response to intervention. Thousand Oaks, CA: Corwin Press. doi:10.4135/9781483329772

Miller F, Cohen D, Chafouleas S, Riley-Tillman T, Welsh M, & Fabiano G (2015). A comparison of measures to screen for social, emotional, and behavioral risk. School Psychology Quarterly, 30, 184–196. doi:10.1037/spq0000085 [PubMed: 25111469]

Mitchell B, Kern L, & Conroy M (2019). Supporting students with emotional or behavioral disorders: State of the field. Behavioral Disorders, 44, 70–84. doi:10.1177%2F0198742918816518

Morrison D (2005). Multivariate analysis of variance. Hoboken, NJ: John Wiley & Sons, Ltd.

National High School Center, National Center on Response to Intervention, and Center on Instruction. (2010). Tiered interventions in high schools: Using preliminary "lessons learned" to guide ongoing discussion. Washington, DC: American Institutes for Research.

Ortuño-Sierra J, Fonseca-Pedrero E, Aritio-Solana R, Velasco A, De Luis E, Schumann G, ... & Bokde A. (2015). New evidence of factor structure and measurement invariance of the SDQ across five European nations. European Child & Adolescent Psychiatry, 24, 1,523–1,534. doi:10.1007/ s00787-015-0729-x

Prewett S, Mellard DF, Deshler D, Allen J, Alexander R, & Stern A (2012). Response to intervention in middle schools: Practices and outcomes. Learning Disabilities Research & Practice, 27, 136–147. doi:10.1111/j.1540-5826.2012.00359.x

Ravens-Sieberer U, Erhart M, Gosch A, & Wille N (2008). Mental health of children and adolescents in 12 European countries-results from the European KIDSCREEN study. Clinical Psychology and Psychotherapy, 15, 154–163. doi:10.1002/cpp.574 [PubMed: 19115436]

Renshaw T (2019). Screening using the SDQ total difficulties Scale: An analog test of three cutoff values. Journal of Psychoeducational Assessment. Advance online publication. doi: 10.1177/0734282918809808

Ruchkin V, Jones S, Vermeiren R, & Schwab-Stone M (2008). The Strengths and Difficulties Questionnaire: The self-report version in American urban and suburban youth. Psychological Assessment, 20, 175ñ182. doi:10.1037/1040-3590.20.2.175

Sansosti F, Telzrow C, & Noltemeyer A (2010). Barriers and facilitators to implementing response to intervention in secondary schools: Qualitative perspectives of school psychologists. School Psychology Forum, 4, 1–21.

Sargisson R, Stanley P, & Hayward A (2016). Multi-informant scores and gender differences on the Strengths and Difficulties Questionnaire for New Zealand children. New Zealand Journal of Psychology (Online), 45, 4–12.

Silverman WK, & Ollendick TH (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. Journal of Clinical Child and Adolescent Psychology, 34, 380–411. doi: 10.1207/s15374424jccp3403pass:[_]2 [PubMed: 16026211]

Smith SR (2007). Making sense of multiple informants in child and adolescent psychopathology: A guide for clinicians. Journal of Psychoeducational Assessment, 25(2), 139–149. doi: 10.1177/0734282906296233

Splett JW, Garzona M, Gibson N, Wojtalewicz D, Raborn A, & Reinke WM (2019). Teacher recognition, concern, and referral of children's internalizing and externalizing behavior problems. School Mental Health, 2, 228–239. doi:10.1007/s12310-018-09303-z

Suldo S, Storey E, O'Brennan L, Shaunessy-Dedrick E, Ferron J, Dedrick R, & Parker J (2019). Identifying high school freshmen with signs of emotional or academic risk: Screening methods appropriate for students in accelerated courses. School Mental Health, 11, 210–227. doi:10.1007/s12310-018-9297-9

Tindal G (2013). Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. ISRNEducation, 2013, 1–29. doi: 10.1155/2013/958530

VanDerHeyden A, Witt J, & Gilbertson D (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of children for special education. Journal of School Psychology, 45(2), 225–256. doi:10.1016/j.jsp.2006.11.004

Vierhaus M, & Lohaus A (2008). Children and parents as informants of emotional and behavioural problems predicting female and male adolescent risk behaviour: A longitudinal cross-informant study. Journal of Youth and Adolescence, 37, 211–224. doi:10.1007/s10964-007-9193-3

Walker H, Severson H & Feil E (2014). Systematic screening for behavior disorders: Administrator's guide, 2nd Edition Eugene, OR: Pacific Northwest Publishing.

Impact Statement:

We found GPA had little utility for capturing secondary students at-risk of emotional and behavioral difficulties, and student-report appeared to capture students with both teacher- and student-reported emotional and behavioral difficulties. Our results demonstrate reliance on teacher-report and extant academic data in universal screening practices in secondary schools may allow adolescents experiencing internalizing and externalizing difficulties to fall through the cracks. Integrating self-report within universal screening practices in secondary schools is therefore an important, and efficient, practice to identify at-risk youth.

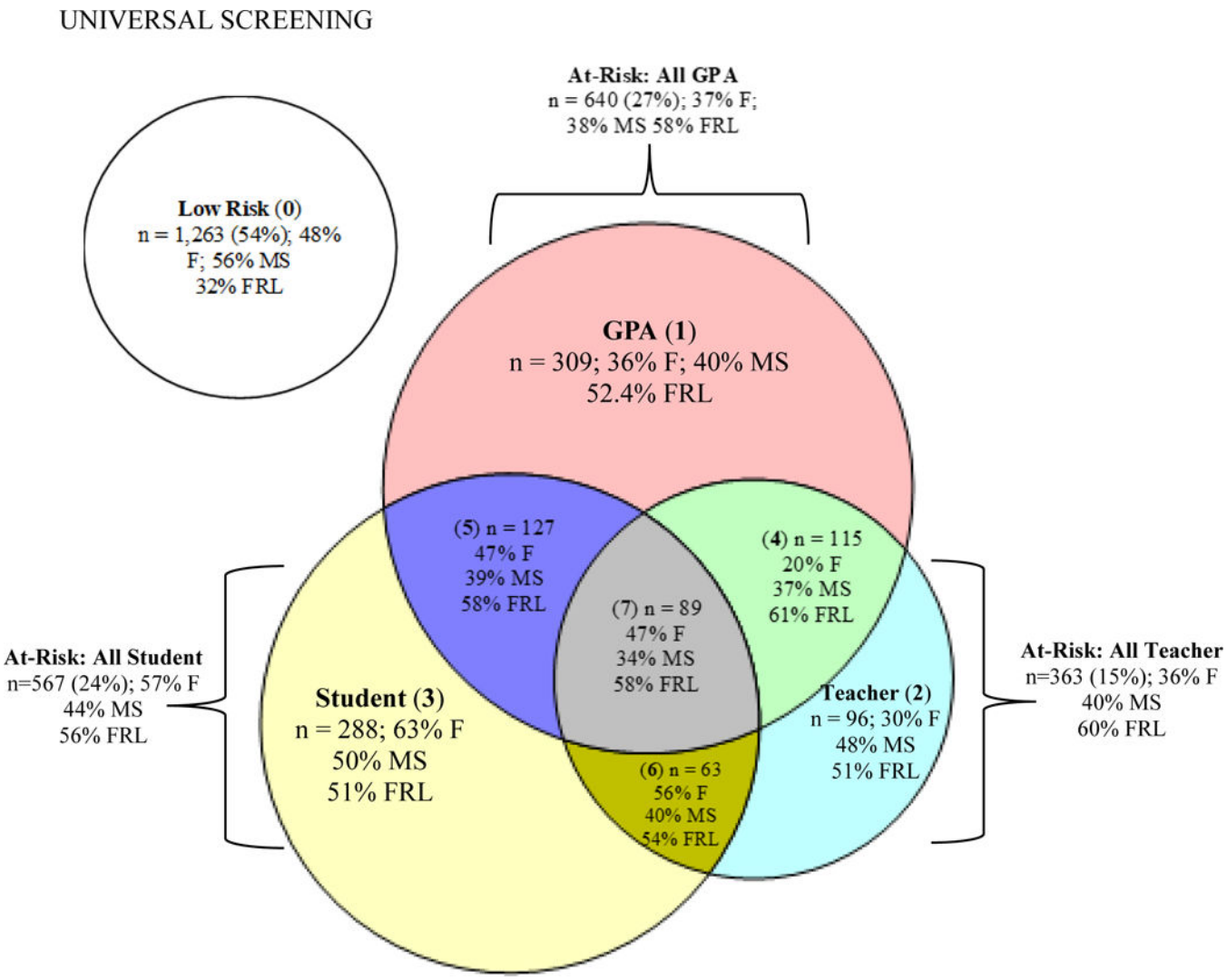

Figure 1.

Descriptive characteristics for students across moderate-risk status. *Note:* N = 2,350. Each circle proportionately represents youth identified by that means of screening (i.e., GPA, teacher-report, student-report). The overlap of two circles represents youth identified by both/all measures. F=Female; MS=Middle School; FRL=qualifies for Free/Reduced Lunches.

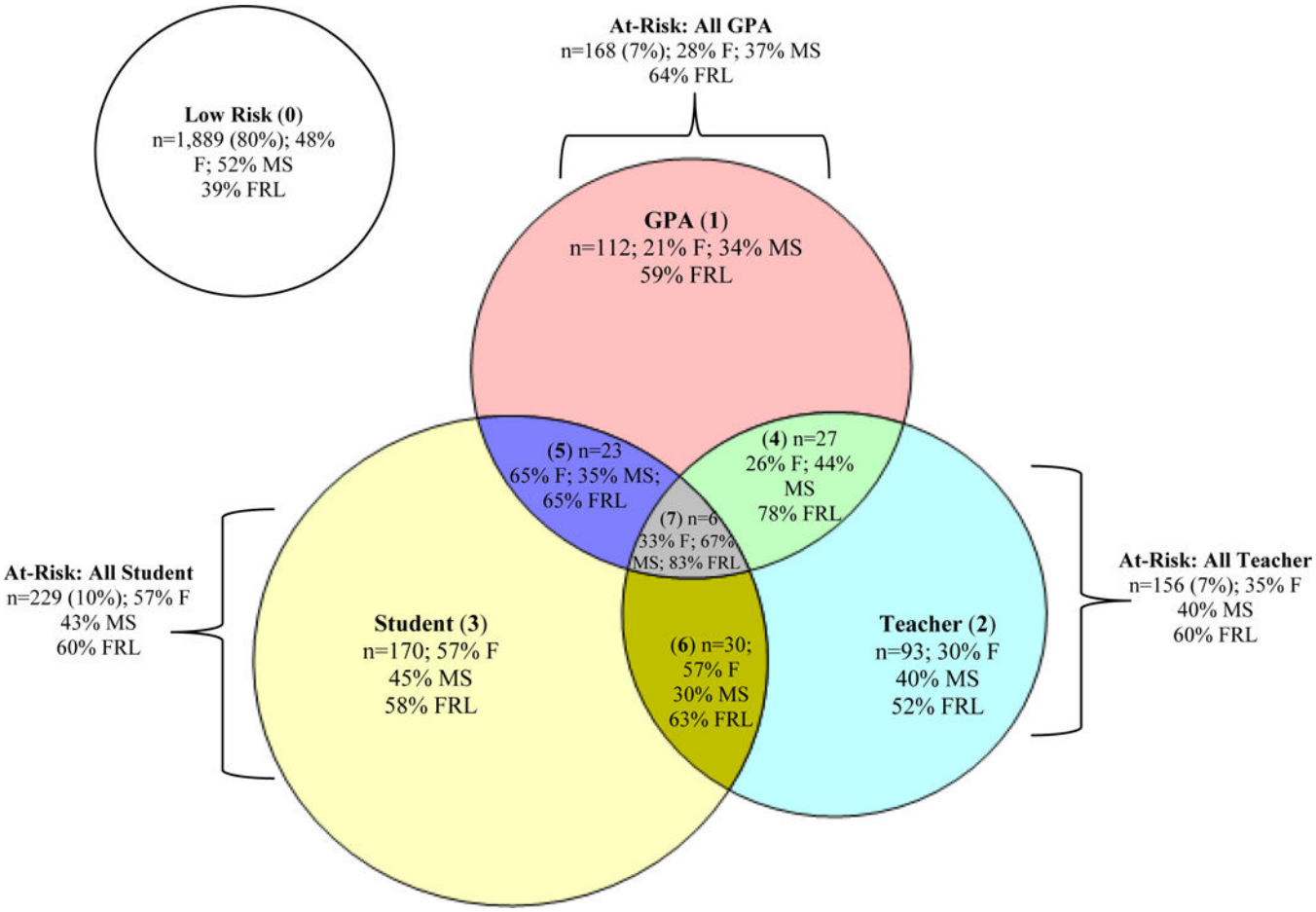

Figure 2.
Descriptive characteristics for students across high-risk status.

Table 1

Indices of agreement across measures

Pearson correlations			Kappa agreement and 95% CI			
			Moderate-risk		High-risk	
	1	2	1	2	1	2
1. GPA						
2. Teacher SDQ	−.45*		.26**[.22, .30]		.15**[.09, .01]	
3. Student SDQ	−.26*	.28*	.14**[.09, .18]	.17**[.13, .21]	.07**[.02, .12]	.12**[.06, .17]

Note.

*$p < .05$.

**$p < .01$

Table 2

Means and 95% CIs of outcome variables across high-risk groups

	High-risk Group							
Subscale and score range	**0** ***M (95% CI)***	**1** ***M (95% CI)***	**2** ***M (95% CI)***	**3** ***M (95% CI)***	**4** ***M (95% CI)***	**5** ***M (95% CI)***	**6** ***M (95% CI)***	**7** ***M (95% CI)***
Teacher Int.	2.15	3.06	8.25	3.18	8.33	3.65	7.60	6.50
(0–20)	(2.04, 2.25)	(2.66, 3.46)	(7.63, 8.87)	(2.75, 3.61)	(7.07, 9.59)	(2.63, 4.67)	(6.37, 8.83)	(4.04, 8.96)
Teacher Ext.	2.59	5.25	11.41	3.48	11.56	4.87	11.70	12.50
(0–20)	(2.46, 2.72)	(4.69, 5.81)	(10.68, 12.14)	(3.03, 3.93)	(10.43, 12.69)	(3.55, 6.19)	(10.55, 12.85)	(9.79, 15.21)
Student Int.	5.21	5.19	5.87	11.65	6.04	10.96	10.90	9.83
(0–20)	(5.07, 5.35)	(4.70, 5.68)	(5.29, 6.45)	(11.23, 12.07)	(5.00, 7.08)	(9.91, 12.01)	(9.91, 11.89)	(7.85, 11.81)
Student Ext.	4.59	6.37	7.04	10.96	7.33	11.87	12.03	12.50
(0–20)	(4.46, 4.72)	(5.84, 6.90)	(6.40, 7.68)	(10.52, 11.40)	(6.51, 8.15)	(10.96, 12.78)	(11.10, 12.96)	(9.79, 15.21)
Teacher Prosocial	7.20	5.05	4.19	6.66	2.96	5.35	3.07	3.06
(0–10)	(7.09, 7.31)	(4.57, 5.53)	(3.73, 4.65)	(6.26, 7.06)	(1.85, 4.07)	(4.28, 6.42)	(2.12, 4.02)	(1.16, 4.96)
Student Prosocial	7.82	7.37	6.65	6.74	6.85	6.87	6.00	7.00
(0–10)	(7.74, 7.90)	(7.02, 7.72)	(6.12, 7.18)	(6.40, 7.08)	(6.09, 7.61)	(6.13, 7.61)	(5.15, 6.85)	(5.10, 8.90)

Note. Groups were coded as depicted in Figure 2.

Table 3

Hedges ‘ g effect sizes. from contrasts comparing different screening measures on outcome variables

High-risk Status							
Comparisons		**Teacher Int.**	**Teacher Ext.**	**Student Int.**	**Student Ext.**	**Teacher Prosocial**	**Student Prosocial**
Group 1	**Group 2**						
GPA (1, 4, 5)	Other at-risk (2, 3, 6)	−0.54	−0.57*	−0.67*	−0.51*	−0.08	0.30
GPA (1, 4, 5)	Low-risk (0)	1.16*	1.61*	0.71*	1.36*	−1.11*	−0.42
Teacher (2, 4, 6)	Other at-risk (1, 3, 5)	1.93*	2.44*	−0.54*	−0.32	−0.92*	−0.26
Teacher (2, 4, 6)	Low-risk (0)	2.39*	3.11*	0.77*	1.45*	−1.53*	−0.71*
Student (3, 5, 6)	Other at-risk (1, 2, 4)	−0.70*	−0.94*	1.77*	1.62*	0.39	−0.22
Student (3, 5, 6)	Low-risk (0)	0.86*	1.13*	1.50*	1.89*	−0.70*	−0.54*
GPA only (1)	Other at-risk (2, 3, 4, 5, 6)	−1.27*	−1.16*	−1.26*	−1.20*	0.24	0.40
GPA only (1)	Low-risk (0)	0.37*	0.92*	−0.01	0.61*	−0.87*	−0.24
Teacher only (2)	Other at-risk (1, 3, 4, 5, 6)	1.25*	1.40*	−0.99*	−0.92*	−0.17	−0.06
Teacher only (2)	Low-risk (0)	2.47*	3.06*	0.21	0.84*	−1.21*	−0.63*
Student only (3)	Other at-risk (1, 2, 4, 5, 6)	−1.21*	−1.90*	1.25*	0.70*	1.02*	−0.01
Student only (3)	Low-risk (0)	0.42*	0.31	2.08*	2.20*	−0.22	−0.58*

Note: The numbers in the left columns indicate the code representing the group used in the contrast.

* indicates $p<.0007$.

Table 4

Percentages of students meeting SDQ subscale high-risk criteria identified by each screening measure

Risk scale	Total *n* at-risk	% Female	GPA % identified	Teacher % identified	Student % identified
Teacher-report					
Emotional problems (≥6)	56	66%	11%	50%	27%
Peer problems (≥5)	170	34%	14%	40%	15%
Conduct problems (≥4)	157	32%	20%	65%	25%
Hyperactivity (≥7)	217	24%	19%	37%	18%
Self-report					
Emotional problems (≥7)	285	73%	5%	7%	44%
Peer problems (≥6)	148	58%	5%	13%	50%
Conduct problems (≥5)	169	31%	13%	23%	57%
Hyperactivity (≥7)	315	46%	11%	14%	44%

Note. Risk status was defined using the total difficulties score from each reporter. Percentages represent the proportion of youth with, for example, teacher-rated at-risk emotional problems identified as at-risk by the GPA cutoff, the teacher SDQ total difficulties score, or the student SDQ total difficulties score. Subscale risk cutoffs are presented in parentheses.